

I hereby certify that this paper is being deposited with the United States Postal Service as Express Mail in an envelope addressed to: BOX PATENT APPLICATION, ASSISTANT COMMISSIONER FOR PATENTS, Washington, D.C. 20231, on this date.

December 12, 2001

Date

Express Mail No. EL846173636US

David Canon

Inventors: Samuel Naffziger; Donald C. Soltis, Jr.

1 A METHOD AND SYSTEM FOR DETECTING DROPPED MICRO-PACKETS

2 BACKGROUND OF THE INVENTION

3 The present invention relates to data transmissions between agents in
4 a network and computer interconnect fabric.

5 Transmissions between agents in a typical network or computer
6 inter-connect fabric are done using "packets" which generally comprise two or
7 more flits or micro-packets that are usually rather small, e.g., 128 bits, to ensure a
8 short transmission time and enable easy handling by very large scale integrated
9 (VLSI) chips along the path. In addition to the data, they contain a small control
10 portion which contains information about the destination locations of the flit and
11 perhaps other information. Dropped flits indicate a failure mode that is not
12 detected by standard cyclic redundancy checking (CRC) or error correction codes
13 (ECC) methods. Parenthetically, such dropped flits can be caused by soft errors in
14 VLSI chips that route the flit to the wrong destination or cause it to be ignored by
15 one of the routers. In this context, soft errors refer to stored information that is
16 lost due to high energy particles resulting from radioactive decay (alpha particles)
17 or gamma rays.

1 Prior art methods of ensuring the reliability of packet transmissions
2 fall into two categories, i.e., flit-level error detection and correction and end to end
3 transmission assurance. Cyclic redundancy check or error correcting codes can
4 check the contents of a flit for errors in transmission, and depending on the code
5 used and the nature of the error, can make corrections. This approach works well
6 to handle error events that operate on the bit level such as electrical noise coupling
7 on the wires used to transmit the data, or random bit flipping in the data portion of
8 the flit.

9 The end to end transmission assurance involves an
10 acknowledgement sequence between the ultimate recipient of a packet and the
11 sending agent. With this method, the receiver of a packet immediately sends an
12 acknowledgement packet to the sender when the complete packet is received. The
13 sending agent must hold a complete copy of each packet sent until the
14 acknowledgement packet is received. This approach works well in handling a
15 large class of errors that can corrupt a packet during its transmission. The cost,
16 however, is high since the sending agent must store all packets that are in flight
17 and must use some sort of time out mechanism to determine if the receiver has not
18 gotten the packet, at which time the sender is required to resend the packet. In
19 addition there is the overhead of the acknowledgement packets consuming extra
20 bandwidth in the network.

21 A need exists to easily detect dropped flits.

22 SUMMARY OF THE INVENTION

23 The present invention comprises a system and a method of providing
24 error detection and correction of transmission of multiple flits between sending
25 and receiving agents connected together in a network or computer interconnect
26 environment that comprises embedding a sequence identifier in each flit prior to

1 transmission, sending each flit to a connected receiving agent, examining the
2 sequence identifiers of each flit being received and requesting the sending agent to
3 resend a flit if the sequence identifier for that flit is determined to be incorrect.

4 In a preferred embodiment of the present invention, the sequence
5 identifier is embedded in the control portion of the flit and comprises a sequence
6 number that is incremented or otherwise changed in a predictable manner, so that
7 the order of flits being received is predicted. If the sequence number for a flit is
8 different than expected, the receiving agent requests that it be resent.

9 DESCRIPTION OF THE DRAWINGS

10 FIGURE 1 is a diagram of a data packet comprising a multiplicity of
11 flits having a control portion and a data portion.

12 FIG. 2 is a diagram of an example of a network with dual processor
13 nodes, and particularly illustrating a packet transmission utilizing multiple hops
14 between two nodes.

15 DETAILED DESCRIPTION

16 The present invention comprises a complimentary error detection
17 and correction approach to the prior art methods of end to end transmission
18 assurance and flit-level error detection and correction, such as cyclic redundancy
19 check and error correcting codes. It is believed to provide a lower cost solution
20 than end to end transmission assurance, but more robust method than flit-level
21 error detection and correction. Failure modes that would not be caught by the flit-
22 level error detection and correction method include errors in VLSI circuitry or
23 wires causing corruption of the control portion of the flit and errors in VLSI
24 circuitry causing the flit to be dropped or lost in its entirety.

1 The system and method of the present invention is intended for use
 2 in the transmission of packets comprised of multiple flits that are transmitted over
 3 one or more hops, i.e., crossing one or more agents, to arrive at a destination
 4 agent. In this regard, an agent is a processor or other VLSI chip such as a memory
 5 controller or input/output (I/O) controller connected in a multiprocessing network
 6 or fabric. As shown in FIG. 1, which diagrammatically illustrates a network with
 7 dual processor nodes and particularly illustrates transmission of from agent 10 to
 8 agent 12, a flit must traverse hops between agents 14, 16, 18 and 20. In the
 9 drawing, agents 16 and 18 are directly connected together.

10 In the present invention, and referring to FIG. 2, a packet 22
 11 typically comprises a plurality of flits that may number from 2 to N with each flit
 12 having a control portion 24 and a data portion 26. The control portion 24 may
 13 have several fields of information such as origination information, destination and
 14 other information (not shown), but importantly to this invention a sequence
 15 identifier that is changed in a predictable manner so that the order in which flits
 16 are sent and received can be determined. While the sequence identifier may be
 17 changed in any predictable manner, the preferred embodiment merely increments
 18 a number by 1 for successive flits. This is carried out by an algorithm which in
 19 pseudo-code comprises:

```

20           if (new flit received {
21           if (flit==data flit && flit != header flit) {
22           Extract sequence number -> s_new;
23           if (s_new != s_old+1 {
24           { signal error to sender:
25           } s_old = s_new;
26           }
27           }
  
```

1 While the foregoing algorithm is used in the preferred embodiment,
2 any predictable incrementing or decrementing operation or digital signature or
3 computation that enables the order of flits to be determined is within the scope of
4 the present invention.

5 To detect dropped flits, the present invention in its preferred
6 embodiment embeds a sequence number in each flit, incremented up from a value
7 that is substantially unique for each packet. As each agent along the transmission
8 path from sender to receiver gets the flit, it checks that this sequence number is the
9 next in line for the packet to which it belongs. If an out of order flit is received,
10 the agent receiving it sends a request for resend to the sending agent, which is not
11 necessarily the original sender.

12 When a sequence number mismatch is detected at the receiving
13 agent, it then signals the sending agent of a failure. This means the sending agent
14 is required to hang on to at least one extra flit in a replay buffer to be able to
15 resend the dropped flit since an error isn't detected until after the subsequent flit is
16 sent. In this regard, whether a copy of the flit is written into a separate replay
17 buffer or merely retained in a memory location is largely a matter of semantics in
18 that one of ordinary skill in the art can manipulate the flit to accomplish the
19 retention and resending of the flit and many alternative types of manipulation is
20 within the scope of the invention. Importantly, the amount of storage required in
21 each agent is quite small since the re-send operation is at an agent-to-agent level,
22 not sender to receiver. In addition, a time out mechanism is avoided since every
23 hop on the transmission path requires either an acknowledgement, or error
24 indication. Such communication can be arranged to consume only a single wire
25 since it is between connected agents in the network.

26 Another benefit of the present invention is that a catastrophic failure
27 of a VLSI chip somewhere in the transmission path will be detected as a missing

1 or incomplete sequence number. This will allow the destination agent to
2 recognize that an error has occurred in this packet and flag the error instead of
3 continuing to consume information with silently corrupted data.

4 From the foregoing, it should be appreciated that a system and
5 method of providing error detection and correction of transmission of multiple flits
6 between sending and receiving agents has been described that has many desirable
7 attributes and advantages compared to known prior art systems. The present
8 invention provides a low cost solution for reliably detecting and correcting errors
9 in transmission of flits that are incapable of being detected and corrected by
10 known prior techniques.

11 While various embodiments of the present invention have been
12 shown and described, it should be understood that other modifications,
13 substitutions and alternatives are apparent to one of ordinary skill in the art. Such
14 modifications, substitutions and alternatives can be made without departing from
15 the spirit and scope of the invention, which should be determined from the
16 appended claims.

17 Various features of the invention are set forth in the following
18 claims.